Dual Process Theories in Behavioral Economics and Neuroeconomics

Grayot, James Daniel

Published in:
Review of Philosophy and Psychology

DOI:
10.1007/s13164-019-00446-9

Publication date:
2020

Review of Philosophy and Psychology (2020) 11:105–136
https://doi.org/10.1007/s13164-019-00446-9

Dual Process Theories in Behavioral Economics and Neuroeconomics: a Critical Review

James D. Grayot [1,2]

Published online: 16 September 2019

Abstract
Despite their popularity, dual process accounts of human reasoning and decision-making have come under intense scrutiny in recent years. Cognitive scientists and philosophers alike have come to question the theoretical foundations of the 'standard view' of dual process theory and have challenged the validity and relevance of evidence in support of it. Moreover, attempts to modify and refine dual process theory in light of these challenges have generated additional concerns about its applicability and refutability as a scientific theory. With these concerns in mind, this paper provides a critical review of dual process theory in economics, focusing on its role as a psychological framework for decision modeling in behavioral economics and neuroeconomics. I argue that the influx of criticisms against dual process theory challenge the descriptive accuracy of dualistic decision models in economics. In fact, the case can be made that the popularity of dual process theory in economics has less to do with the empirical success of dualistic decision models, and more to do with the convenience that the dual process narrative provides economists looking to explain-away decision anomalies. This leaves behavioral economists and neuroeconomists with something of a dilemma: either they stick to their purported ambitions to give a realistic description of human decision-making and give up the narrative, or they revise and restate their scientific ambitions.

1 Introduction

Dual process theory (DPT) has been playing a prominent role in both the cognitive and behavioral sciences. The standard view of DPT —also known as the "received view" (Evans and Stanovich 2013b; cf. Mugg 2016)—suggests that different aspects of human cognition, such as reasoning, judgment, and decision-making, can be

✉ James D. Grayot
james.grayot@gmail.com

[1] Tilburg Center for Logic, Ethics, and Philosophy of Science (TiLPS), Tilburg University, Warandelaan 2, 5037 AB Tilburg, Netherlands

[2] Faculty of Philosophy, University of Groningen, 9712 CP Groningen, Netherlands

categorized according to and/or understood as the result of these two mental processing types. "Higher" mental processes are depicted as *slow*, *controlled*, *reflective*, *serial*, *rule-based*, *effortful*, and *conscious*, and are associated with energy-intensive cognitive tasks like deductive reasoning and hypothetical thinking. This category of mental processing is commonly referred to as "System 2" or "Type 2". By contrast, "lower" mental processes are depicted as *fast*, *reactive*, *automatic*, *intuitive*, *heuristic*, *associative*, and *unconscious* (or *preconscious*), and are associated with perceptual and affective operations like attentional cueing and motor-response preparation. This category of mental processing is commonly referred to as "System 1" or "Type 1" processing.[1]

Over the last few decades, economics has sought increasing support from cognitive psychology and neuroscience to improve decision models, often employing the concepts and rhetoric of DPT. The first, and perhaps best-known example of this is the study of judgment and decision-making under risk and uncertainty in the Heuristics and Biases tradition (Tversky and Kahneman 1973, 1974; Kahneman et al. 1982; Kahneman and Frederick 2002, 2005). The second is the refinement of intrapersonal and intertemporal choice models for studying time preferences and self-control problems (Loewenstein 1996, 2000; Bénabou and Tirole 2002; Bernheim and Rangel 2004; Benhabib and Bisin 2005; Loewenstein and O'Donoghue 2005; Fudenberg & Levine 2006). Unlike neoclassical decision models that presume agents to be faultless utility-maximizers and which often resort to ad hoc explanations to justify deviations from expected utility theory, DPT has provided behavioral economists and neuroeconomists with psychologically plausible foundations to represent the internal dynamics of decision-making (Grayot 2019).

DPT has come under intense scrutiny during the last two decades. Cognitive scientists and philosophers alike have come to question the theoretical foundations of the standard view of dual process theory and have challenged the validity and relevance of evidence in support of it (Gigerenzer and Regier 1996; Osman 2004; Keren and Schul 2009; Kruglanski and Gigerenzer 2011). Moreover, attempts to modify and refine dual process theory in light of these challenges have generated additional concerns about its applicability and refutability as a scientific theory (Keren 2013; Mugg 2016; Pennycook 2017; Bonnefon 2018). This should raise concerns for behavioral economists and neuroeconomists who see DPT as providing more than merely psychologically plausible foundations for their models (cf. Angner 2019). In particular, it raises the possibility that dualistic decision models in economics are not as descriptively accurate or reliable as economists presume them to be. In fact, the case can be made that the popularity of DPT in economics has less to do with the empirical success of dualistic decision models, and more to do with the convenience that the dual process narrative provides economists looking to sort out and explain-away decision anomalies (cf. Grüne-Yanoff 2017; Angner 2019). I argue that the growing number of criticisms against DPT leaves economists with something of a dilemma: either they stick to their purported ambitions to give a realistic description of human decision-making and give

[1] For classic texts on dual processing, see Shiffrin and Schneider (1977), Evans (1989), Epstein (1994), Sloman (1996), Stanovich and West (2000), Lieberman et al. (2002). For recent developments and modifications, see Stanovich (1999, 2004, 2009, 2011), Evans (2006; 2008; 2009a, b; 2011; 2012), Evans & Frankish (2009), Evans and Stanovich (2013a, b), and De Neys (2017).

up on DPT, or they stick to DPT and revise and restate their scientific ambitions. To motivate this dilemma, this paper raises two challenges:

The first challenge pertains to how dualistic decision models in economics represent choice as the *outcome* of dual processes and/or systems. This challenge is twofold: (1) In the case of neuroeconomics, DPT has been employed to illustrate how the brain evaluates prospects and how it executes decisions; this has led researchers to identify specialized mechanisms and sub-systems in the brain with different mental functions. However, because DPT is not per se a mechanistic theory, dualistic decision models in neuroeconomics may exaggerate or distort the roles that mechanisms or sub-systems play in the valuation of prospects and the execution of decisions. On the one hand, this may lead to confusion about what, exactly, DPT is meant to represent; on the other hand, it may render DPT descriptively redundant. (2) However, DPT is far more prevalent in behavioral economics, with the majority of dualistic decision models taking a broadly functionalist approach toward the description of the mental processes that underpin decisions. While this is consistent with the standard view of DPT, it also means that dualistic decision models in behavioral economics cannot explain in a non-question-begging way how choice emerges from the interaction of dual processes and/or systems—that is, not without a supplemental story. This leaves behavioral economists with two options: (i) bite the bullet and leave the black box of the mind closed; (ii) use formal constructs from economics to work out the details of how possibly dual processes and/or systems interact and produce choice. The problem with the latter option is that the explanatory value of dualistic decision models becomes hostage to the explanatory value of DPT in general. As we'll see, this explanatory value is suspect.

The second challenge addresses a more fundamental question with regard to DPT, namely, how does it distinguish rational from irrational decision processes? Despite the commonly held belief that System 2 (Type 2) is necessary for, or synonymous with, rational decision-making and System 1 (Type 1) is either irrational or non-rational, there are both theoretical and empirical considerations to suggest that this dichotomy is not tenable, and perhaps in need of serious revision. Hence, economists are overly optimistic if they think that DPT, construed as a functional framework, provides dualistic decision models with the normative foundations they require. While this does not warrant abandoning a dual process view in toto, it certainly calls into question whether in vogue economic models can reliably predict and/or explain decision phenomena outside the lab.

Lastly, it should be made clear that the aim of this paper is not to disparage behavioral economists and neuroeconomists for their attempts to develop more realistic decision models. Quite the opposite. The aim is to address open questions which concern DPT as a psychological framework and, to that end, to better understand its limitations in the behavioral sciences.

This paper is structured as follows: Section 2 surveys the explicit and implicit influences of DPT upon behavioral economics and differentiates two styles of dualistic modeling. Section 3 provides an overview of the theoretical and empirical criticisms of DPT from the perspective of cognitive science and philosophical psychology: it discusses the main differences between system-based and type-based theories of information processing. Section 4 considers the challenges for dualistic decision models in behavioral economics and neuroeconomics, arguing for the limitations of each style with regard to how they employ DPT. Section 5 then considers whether DPT

can sustain the normative foundations of economic models given that neither system-based nor type-based interpretations can cleanly distinguish rational from irrational decision processes. Section 6 concludes.

2 How Dual Process Theory Made its Way into Economics

Behavioral economics has earned a reputation for being psychologically realistic and for providing new insights into the hidden processes that govern decision-making. This is due in large part to mass market publications, such as Dan Ariely's *Predictably Irrational* (2008), Richard Thaler's *Nudge* (co-authored with Cass Sunstein 2008), and Daniel Kahneman's *Thinking, Fast & Slow* (2011), which portray behavioral economics as an exciting new discipline that has the potential to unlock the mysteries of the human mind. This image has been further reinforced by Kahneman and Thaler each receiving Nobel Prizes for their contributions to economics. This reputation has, in part, strengthened economists' faith in the predictive and explanatory power of DPT. In this section, I make the case that at least two research programs in behavioral economics have been deeply inspired by DPT (or core features of it). The first, and perhaps best well-known, is the study of judgment and decision-making under risk and uncertainty in the Heuristics and Biases tradition. The second is the refinement of intrapersonal and intertemporal choice models for the study of time preferences and impulse-control.[2]

2.1 Explicit and Implicit Examples of Dual Processing in Behavioral Economics

In order to appreciate how the dual process narrative has permeated economic decision research, we need to look at the historical episodes that brought economics and psychology into close proximity. The development of the Heuristics and Biases research program is such an episode.

It is now well understood that one of the major threads in behavioral economics, namely the study of judgment and decision-making under risk and uncertainty, has its origins in the behavioral decision research of Tversky and Kahneman (1973, 1974) and Kahneman et al. (1982)—for historical overviews, see Sent (2004), Heukelom (2014), Angner (2019). The goal of this research was, primarily, to understand why individuals tend to make mistakes when forming probabilistic judgments about choices and how to predict when cognitive load may compromise one's reasoning ability. The principal discovery in this research is that individuals often resort to shortcuts and other rules-of-thumb to facilitate decision-making; sometimes these shortcuts are helpful, but often they can be biased in ways that undermine one's reflective or computational abilities. Original studies posited three primary heuristics—*accessibility*, *representativeness*, and *anchoring* (Tversky and Kahneman 1973; Kahneman et al. 1982)—as the likely cause of reasoning errors; and it was clear that Tversky & Kahneman saw these heuristics as the result of differential mental processing, consistent with research in social and

[2] Although DPT has had a major influence on certain branches of behavioral economics, I'm not here claiming that *all* behavioral economic models have been influenced by it. For instance, there is no explicit mention of dual processing in either Tversky & Kahneman (1992) or Wakker (2010) regarding the psychological components of cumulative prospect theory. Likewise, there is no mention of dual processing in Laibson (1997) regarding the causes of the quasi-hyperbolic discounting function.

cognitive psychology of the 1970's (cf. Shiffrin and Schneider 1977; Nisbett and Ross 1980).[3]

Subsequent research on psychological applications of the heuristic and biases research have reinforced the idea that human decision-making involves the interplay of two types of mental processes (Sloman 1996; Chaiken and Trope 1999; Gilbert 1999, 2002). The most vivid demonstrations of the mind's two systems are explicated in Kahneman & Fredrick (2002; Kahneman and Frederick 2005; Kahneman and Frederick 2007) and Kahneman (2003a, b, 2011) who adopt the terminology (from Stanovich and West 2000) of "System 1" and "System 2". For all intents and purposes, this understanding of System 1 and System 2 bears all the core characteristics of other system-based interpretations of DPT (see next section). For instance, Kahneman and Frederick (2002, 2005) claim that System 1 corresponds to "intuitions", which are directly informed by the perceptual system, whereas System 2 corresponds to "reflective judgments", which cautiously assess intuitions and formulate responses to them:

> In the particular dual-process model we assume, system 1 quickly proposes intuitive answers to judgment problems as they arise, and system 2 monitors the quality of these proposals, which it may endorse, correct, or override …We assume system 1 and system 2 can be active concurrently, that automatic and controlled cognitive operations compete for the control of overt responses, and that deliberate judgments are likely to remain anchored on initial impressions. (Kahneman and Frederick 2005, p. 267)

While Kahneman is perhaps the most vocal proponent of DPT, the dual process narrative has received further support and attention in Thaler's research on libertarian paternalism (Thaler and Sunstein 2003, 2008; Thaler et al. 2012). Thaler appeals to the differential processing abilities of ordinary decision-makers to justify *nudging*, i.e. improving choice architectures through subtle framing interventions. The psychological evidence in support of nudging comes directly from classic texts in the DPT literature (e.g., Chaiken and Trope 1999; Lieberman et al. 2002) as well as from the heuristics and biases research itself (e.g., Tversky and Kahneman 2000; Gilovich et al. 2002). For the sake of space, I will not review Thaler's conception of DPT as it essentially mirrors that above.[4]

Of course, not all dualistic economic models are as explicit about their commitment to DPT as those in the Heuristics and Biases tradition and nudge research. Another important episode in behavioral economics in which DPT has played a recurring theme is the study of intrapersonal and intertemporal choice. Grayot (2019) analyzes how researchers have attempted to integrate dual process and dual system theories of mental processing with multiple-self economic modeling techniques, giving rise to a new breed of psychologically sophisticated "multiple-agent" models. The most cited

[3] Kahneman (2011) explains in retrospect that prospect theory was, in essence, a formalization of the dual-process account developed in Heuristics and Biases research: "Although Amos and I were not working with the two-systems model of the mind, it's clear now that there are three cognitive features at the heart of prospect theory. They play an essential role in the evaluation of financial outcomes and are common to many automatic processes of perception, judgment, and emotion" (2011, p. 273).

[4] For a critical reflection on Thaler's legacy in behavioral economics, especially with regard to his attempts to make economics psychologically more realistic, see Grüne-Yanoff (2017).

examples of this sort of model include Bénabou and Tirole (2002), Bernheim and Rangel (2004), Benhabib and Bisin (2005), Loewenstein and O'Donoghue (2005) and Fudenberg and Levine (2006), to name a few. In particular, multiple-agent models of intrapersonal and intertemporal choice have sought to characterize the "contradictory tendencies of temporally distinct selves by investigating how controlled and automatic processes influence choice behaviors over time" (Grayot 2019, p. 4). In some instances, the internal dynamic between intrapersonal selves is interpreted to establish the limitations on the decision-maker's ability to exhibit self-control; in other instances, the conflict between desires to consume now or later is interpreted as a 'trade-off' which is determined by the activation of distinct cognitive (or neural) processes.

These less-explicit instances of DPT in behavioral economics reveal that the dual process narrative can serve different modeling purposes: whereas the former approaches, i.e. Bénabou & Tirole (2002, 2003), Fudenberg & Levine (2006), adopt a dual-*self* model in which the decision-maker is temporally divided, the latter, i.e. Benhabib and Bisin (2005), Loewenstein and O'Donoghue (2005), start with a conception of the decision-maker who is already psychologically divided into dual-*systems*; the model then seeks to understand how both static and dynamic choice phenomena result from interactions between systems.

The table below (Table 1) provides a list of different variations of dualistic decision models in economics; though they employ different naming conventions, the same logic applies to all, which is that dualistic structures give rise to internal conflict and produce different sorts of choice phenomena.

One will notice that the references cited in the table cover a range of modeling techniques, not all of which are typically associated with behavioral economics (I return to this point shortly). Nevertheless, they share two things in common which speaks to the influence of DPT on behavioral economics and neighboring subdisciplines: First, each adopts an overtly dualistic rhetoric (even if this has no associated theoretical or empirical commitments). Second, each relies on a codification of the dualistic rhetoric which results in a constrained optimization model of choice (e.g., principal-agent model or limited information game). In fact, it's not hard to see how psychologically sophisticated multiple-agent models in behavioral economics and neuroeconomics emerged

Table 1 Different variations of dualistic decision models

Dualistic label	Reference
"*Cold*" vs. "*hot*" *modes*	Loewenstein 1996; Bénabou and Tirole 2002; Bernheim and Rangel 2004, 2007; Ainslie 2001, 2005; Soman et al. 2005
Cognition vs. *emotion*	Sanfey et al. 2003
Deliberative system vs. *affective system*	Loewenstein and O'Donoghue 2005
Cognitive system vs. *affective system*	Camerer et al. 2005
Reflective vs. *impulsive system*	McClure et al. 2004, 2007; Strack and Deutsch 2004; Fudenberg & Levine 2006;
Controlled vs. *automatic processing*	Loewenstein 1996, 1999; Camerer et al. 2005; Benhabib and Bisin 2005
Executive control vs. *conflict monitoring*.....	Benhabib and Bisin 2005; Brocas and Carrillo 2008, 2014

from prototypical "planner–doer" and "multiple-self" models (Thaler and Shefrin 1981; Schelling 1984; Shefrin and Thaler 1988; Ainslie 1992; Ainslie and Haslam 1992). The particular arrangement of the planner-doer model indicates that the planner *self* is not just farsighted, but that it has rational authority, which is realized when it exerts control over the doer *self*. This is a clear indication that economists have sought to resolve the problem of dynamic inconsistency by invoking asymmetrical reasoning processes, which may supervene on real psychological or physiological processes. So, while the case can be made that traditional planner-doer and multiple-self models presuppose a kind of differential mental processing, the behavioral economic models listed in the table above place more emphasis on uncovering the cognitive and neurobiological basis of willpower and self-control.

One possible explanation for why intrapersonal and intertemporal choice models have become so diverse may have something to do with how the literature on time preferences and choice inconsistency has developed. For example, Loewenstein (1996, 2000) and Metcalfe and Mischel (1999) are among the most frequently cited in the behavioral economics literature with regard to the psychological underpinnings of inconsistent preferences. Loewenstein and Metcalfe & Mischel both emphasize the significance of visceral factors (e.g., cravings, sexual arousal, pain) and heightened emotional states (e.g., "hot" states) as responsible for impulsive or unreflective choices. Thus, the hot-cold heuristic has percolated throughout behavioral economics and has been adopted by many as the default psychological model for the study of self-control problems. In this way, the hot-cold heuristic serves nearly the same function that DPT does—in fact, the only discernable difference between the hot-cold heuristic and DPT is the disciplines from which they emerge: The hot-cold heuristic is more closely associated with the psychometrics of willpower and delay-gratification, which has close links to utility theory as it is employed in mathematical psychology.

2.2 The Increasing Popularity of Dualistic Decision Models in Economics

How do we reconcile the increasing popularity of dualistic models in behavioral economics with the controversies surrounding DPT? One answer is that economists simply aren't familiar with the debates in cognitive science and philosophical psychology about DPT, and thus shouldn't be expected to know each and every criticism against it (especially if those criticisms are still up for debate). However, it could also be the case that behavioral economists think these criticisms do not apply to them, perhaps because their descriptive aims are not so tied up in the explanatory power of DPT. As we saw above, while some dualistic decision models are explicit about their commitments to DPT (such as in the Heuristics and Biases tradition), many are not explicit about this, either because they don't recognize the historical links, or because they interpret dual processing through another disciplinary frame. For this reason, it's difficult to know whether the controversies and criticisms of DPT carry over and have implications for behavioral economics and neuroeconomics. To answer this question, we need say a bit more about the possible roles that DPT plays in dualistic models.

Some dualistic decision models make use of neuroscientific evidence and appeal to executive mechanisms and/or specialized sub-systems in the brain; these mechanisms and sub-systems are believed to play a pivotal role in the execution of decisions and thus are thought to be highly relevant to the analysis of rational choice. In these cases,

which characterize its role in neuroeconomics, DPT not only provides empirical support for dualistic modeling, but it possibly opens the proverbial black box by pinpointing physiological structures that have been left out of economic analysis in the past (Camerer et al. 2005; Camerer 2007; cf. Bernheim 2009). Yet, the proportion of dualistic models which rely on neuroscientific evidence is small. By contrast, the majority of dualistic decision models in behavioral economics can be understood as taking a purely functional (non-reductive) approach to the analysis of decisions. In these cases, DPT could be thought to provide a menu of mental processing types, which would allow behavioral economists to make broader inferences about the etiology of choice phenomena.

So, where does this leave us? As I stated in the introduction, the controversies and criticisms brought against DPT give rise to a dilemma, one which requires behavioral economists and neuroeconomists to reflect on their psychological commitments and possibly revise their scientific ambitions. But it's also possible that DPT could play different descriptive roles with regard to dualistic decision modeling.

In the next two sections I show how the theoretical and empirical inadequacies of DPT are directly relevant to economics: in section 3, I survey six essential criticisms that have been raised against DPT from the perspective of cognitive science and philosophical psychology; in section 4, I then examine how these criticisms impact the descriptive accuracy of decision models, starting with specific instances in neuroeconomics, and then segueing into behavioral economics for a more general assessment.

3 Recent Developments in Dual Process Research

3.1 Taking a Closer Look at System 1 and System 2

Nearly all versions of DPT subscribe to the same basic idea, which is that human minds rely on distinct types of mental processing to accomplish different tasks in daily life.[5] It's widely believed that these processes evolved for specific purposes and are designed and attuned to respond to features of the external environment. As previously mentioned, the standard view distinguishes mental processes that are fast, reactive, automatic, intuitive, heuristic, associative, and preconscious from mental processes that are slow, controlled, reflective, serial, rule-based, and conscious. Mugg (2016) refers to this as the "standard menu" of mental processes; I will refer to it as the standard menu throughout this paper.

On the standard menu, processes come in two types. As previously mentioned, the former is believed to be evolutionarily old and directly linked to autonomic functions, such as 'fight or flight' responses and stimulus-bound perceptions. The latter set of processes are believed to have evolved more recently and aid in higher cognitive functionings that draw upon working memory and require sustained effort and attention (Evans and Over 1996; Stanovich 2004). Classic experiments, such as the Wason Selection Task (Evans and Wason 1976; Evans 1989) and Stroop Effect test (Stroop

[5] For overviews of DPT's applications and interpretations across psychology See Evans (2006, p. 208) and Pennycook (2017).

1935; Osman 2004), demonstrate how reasoning errors and biases depend largely on the allocation of cognitive resources, which are determined by the automaticity of information processing protocols. Certain processes—typically those associated with older evolutionary structures—are easily primed and often trigger responses before an individual can, say, consult a rule or deliberate about a problem. In the domain of reasoning and decision-making, the effects of automatic and rapid processing can be observed through misapplied decision heuristics and faulty reasoning, as well as computational errors. These experiments are believed to give credence to the validity of DPT.

If this sounds somewhat vague, however, it's because DPT *is* vague. The constellation of theories that make up DPT are better thought of as a generic framework than a unified theory (Evans and Stanovich 2013a, b). Hence, a well-known defect of the standard view is that it's not obvious what distinguishes different mental processes from one another, aside from the labels the theory ascribes to their functional roles. Moreover, it's difficult to know whether token theories utilizing the standard menu of mental processes refer to the same thing. (This, it should be noted, is a problem for functional explanations in psychology in general—Levin 2017). Consequently, the standard view of DPT does not actually provide an account of how reasoning tasks are accomplished, and decisions made; what it provides is a generic theory about the potential origins of reasoning and decision errors. This, it would seem, is a major deficiency for the theory: if it cannot explain how the mind inhibits or overrides bad judgments that are generated by rapid or automatic mental processes, then what is the point of making the distinction to begin with? After all, we don't always submit to our biases—we are often able to restrain gut-reactions and to recognize hasty errors for what they are. But the fact that we frequently make reasoning errors under conditions of risk and uncertainty indicates that much of our mental processing is not under conscious control.

Theorists have responded to this issue by positing separate modes of processing, parsing the standard menu into discrete systems. These are most commonly known as *System 1* and *System 2* (Stanovich 1999; Kahneman 2003a, b, 2011; Kahneman and Frederick 2005); though, some authors have opted for less neutral terminology, referring to them as the *Heuristic System* and *Analytic System*, respectively (Stanovich 2004; Evans 2006; cf. Evans and Stanovich 2013a). For an extensive overview of the different clusters of attributes that are said to belong to System 1 and System 2, see Evans (2006, 2008).

While the idea of separate cognitive systems has helped to synthesize many token theories in the DPT literature, thereby alleviating some of the worry about terminological discrepancies, perhaps the most important—and arguably the most controversial—aspect of the System 1 / System 2 framework is the way the two systems are believed to *interact*. One interpretation is that the systems are arranged sequentially: System 1 operates autonomously, with System 2 monitoring and intervening when it has the power, i.e. energy resources, to do so. This is known as the "default-interventionist" model. Another interpretation is that the two systems are arranged in parallel and must compete for control over our behavior. This "parallel-competitive" model is appealing given new evidence about the distribution of brain processes (Sinayev 2016; Lurquin and Miyake 2017; Pennycook 2017). Yet, the consensus among many researchers, at

least in the areas of reasoning and decision-making, is that this latter view is untenable (cf. Evans 2008; Keren and Schul 2009).[6]

The reason why many seem to resist the parallel-competitive model of system interaction is because it requires a more complex explanation about how System 1 and System 2 co-operate and reconcile conflict. For proponents of the default-interventionist model, there is no conflict per se; System 1 operates autonomously and System 2 either intervenes or it doesn't. But on the parallel-competitive model, both systems are thought to generate responses to input, and although System 2 can override System 1, the associative force of System 1's responses may block System 2's attempts to intervene. While it is still a debated issue which description better approximates the interaction of System 1 and System 2, proponents of both agree that System 2 could not operate without System 1 because the higher cognitive functionings of System 2 *depend* on information received by System 1 (cf. Evans and Stanovich 2013a). It's important to keep in mind here that "parallel" has different interpretations and can refer to or range over different operations within a system. In this instance, parallel is meant to encompass the operations of both System 1 and System 2, meaning that each system is designed to respond to unique inputs and does not overlap or share functional characteristics with the other—in a word, they operate autonomously. This can be contrasted from other instances of parallel and distributed processing which occur at the intra-system level. For instance, some who endorse the default-interventionist model readily acknowledge that within System 1 there may exist many autonomous sub-systems which operate in parallel (isolated from one another, but not isolated from the higher reflective processes of System 2). This is the basis of Stanovich's concept of *The Autonomous Set of Systems* which comprise the Heuristic System (Stanovich 2004, 2011).

With this in mind, some have speculated that if System 1 and System 2 operated independently of one another (which the parallel-competitive model suggests), then it seems the only way they could meaningfully interact and compete for control over our behavior is by way of a third system, which has access to the inputs and processes of both System 1 and System 2 (Stanovich 2011; Varga and Hamburger 2014). Indeed, there is growing neuroscientific support for the existence of executive control functions in the brain, and some supporters of the system-based interpretation ascribe this function to System 2. But the range of processes that this executive function has control over appears to be limited (cf. Pennycook 2017); moreover, if executive control were a feature of System 2, this would indicate that System 1 and System 2 are not isolated from another (otherwise System 2 couldn't perform its role as executor—Keren and Schul 2009). For this reason, the default-interventionist model is, at least in the domain of decision-making, the more plausible of the two models of system interaction.

Despite the proliferation of DPTs that use the terminology of System 1 and System 2, there are several outstanding criticisms of the systems-based interpretation—many of

[6] Arguably, one could make the case for a third arrangement, wherein systems process information simultaneously (in parallel), but they are allowed to influence each other in complex, feedback interactions (Sinayev 2016). For the sake of space, I will not consider this interpretation of system interaction here for it is not common in behavioral economic literature. Moreover, for reasons that will become evident in section 4, I suspect that this arrangement would not be conducive to behavioral economic modeling since the utility function of one system would presumably change upon feedback interactions with the system it is in "competition" with. For alternative approaches to modeling system interactions in neuroeconomics, see Krajbich and Dean (2015) and Konovalov and Krajbich (2019).

which have not received due attention outside the cognitive sciences. Consider the following three criticisms:

Criticism 1: Systems are not discrete. It is reasonable to think that System 1 and System 2 roughly correspond to neuroanatomical differences in the brain; and, it's been suggested by many (with the support of brain imaging software) that some functions of System 1 and System 2 can be correlated with domain-specific modules and/or neural circuits (Mars et al. 2011; Botvinick and Cohen 2014; Botvinick and Braver 2015). However, there is not sufficient evidence to warrant *identifying* System 1 and/or System 2 with any fixed neural architecture (Osman 2004; Keren 2013). Rather, evidence indicates that many processes associated with both systems "crosscut" each other for the sake of executing different functions (Mugg 2016; Keren and Schul 2009; Evans and Stanovich 2013a). This has two important consequences for the descriptive accuracy of DPT: firstly, it indicates that System 1 and System 2 are not discrete—in fact, they may share or utilize similar neural pathways for the completion of dissimilar tasks. This is not so surprising when one considers that the standard menu of processes is characterized according to the *functions* of System 1 and System 2, and these functions may be multiply realized depending on the task at hand or the circumstances surrounding a task. Secondly, as Keren and Schul (2009) have pointed out, the contrastive nature of System 1 and System 2 really is a matter of degree, as mental processing occurs along a continuum (e.g., the dividing line between "controlled" and "automatic" processing, as for many other mental processes, is fuzzy and indistinct, and may be different for different individuals).

Criticism 2: Intersystem interactions are underdetermined by evidence. As briefly described above, the issue as to whether systems are arranged in a sequential or parallel fashion is very much a contingent (and debated) matter: it depends entirely on how one defines the concept of a cognitive system and how this is fleshed out in terms of its functional characteristics. Because it is not agreed upon what the appropriate neuroanatomical correlates of System 1 and System 2 are, the story of their interaction is mired in theoretical and terminological disputes. Although most researchers prefer to believe that System 1 and System 2 are arranged sequentially, there simply isn't sufficient empirical evidence to validate either the default-interventionist model or the parallel-competitive model of system interaction. Recent meta-analyses and replication studies indicate that neither model is singularly equipped to predict and explain how individuals reason and make decisions (Sinayev 2016; Lurquin and Miyake 2017; cf. Pennycook 2017). The evidence and counter-evidence to support both models could be interpreted as a fundamental flaw in the theory itself.

Criticism 3: Evidence for dual systems is limited to laboratory settings. Finally, there is growing consensus among critics that system-based interpretations of DPT is predictive only insofar as it predicts behavior in highly controlled, laboratory settings (Keren 2013; Buturovic and Tasic 2015). On the one hand, it has not been proven that either system is solely responsible for reasoning errors (and, as we'll see in section 5, there are important reasons to believe that System 2 is not only capable of reasoning errors, but that making occasional errors is part of its function). On the other hand, the case has been repeatedly made that proponents of the system-based interpretation (see e.g., Kahneman and Frederick) presupposes

norms of rationality that based on rules of logic and probability theory. This emphasis on testing peoples' abilities to solve puzzles and perform computational tasks in artificial conditions says little about their day-to-day reasoning abilities (Gigerenzer and Regier 1996; Gigerenzer and Brighton 2009; Kruglanski and Gigerenzer 2011). It has been further argued that the system-based interpretation of DPT relies on biased results, and that experimenters are selective in their reporting of evidence (Gigerenzer 2015). It goes without saying that the above criticisms have generated much controversy.[7]

3.2 Why the Type 1 / Type 2 Distinction doesn't Escape Criticism

It could be argued that the above criticisms, while valid, do not undermine the theoretical significance of DPT; rather, they merely demonstrate the limitations of particular models and particular applications of it. For instance, Evans and Stanovich (2013a, b) now acknowledge that the system-based interpretation of DPT has many deficiencies. Though, they maintain that such criticisms also betray a confusion by critics between theory and meta-theory, and they maintain that DPT—construed as a *meta-theory*—has not been, or rather, cannot be refuted (Evans and Stanovich 2013b; Pennycook 2017). What Evans & Stanovich mean by "meta-theory" is not altogether clear. They claim that, "Broad frameworks, like dual-process theory, have a very important role to play in psychology, and there are numerous examples of research programs organized within and around such frameworks… What we can expect at this level is general principles, coherence, plausibility, and the potential to generate more specific models and the experiments to test them" (2013b, p. 263). As such, Evans and Stanovich have since abandoned the system-based interpretation, arguing instead that DPT is most plausible if mental processes are organized by a single dichotomy, namely, their *type*; hence they adopt the terminology Type 1 and Type 2 to distinguish the processes which they formerly attributed to the Heuristic System and Analytic System, respectively.

While proponents of this new type-based interpretation are cautious not to overstate the discreteness of Type 1 and Type 2 processes, they utilize many of the same attributes to differentiate the two: Type 1 consists of autonomous processes that are automatic and not under an individual's conscious control, whereas Type 2 consists of reflective processes that are typically associated with activities like language use, mental simulation, and complex problem solving. However, what really sets Type 1 / Type 2 apart from System 1 / System 2 is the role of working memory in higher-order functionings (Evans and Stanovich 2013a).

Stanovich (2009) further modifies the type-based interpretation, positing that in addition to the autonomous set of systems that make up Type 1, Type 2 processes can be further bifurcated into distinct stages: the first stage involves what he calls "algorithmic" processing, which initiates many of the monitoring and executive functions that are associated with the Analytic System. It is only after algorithmic

[7] Although Evans and Stanovich have elected not to use the terms "System 1" and "System 2" to characterize DPT, they acknowledge that most of the criticisms presented above apply to their own conception of Heuristic System and Analytic System.

processing that the second stage of Type 2 processing is engaged, where genuine "reflective" processing takes place (Stanovich 2011). The algorithmic stage of mental processing is an important innovation in this model, as it is intended to mediate between the autonomous set of systems while effectively priming information for conscious manipulation. The significance of positing an algorithmic "level" is that it is thought to account for discrepancies in the application of DPT, such as individual differences in intelligence and cognitive ability. Nevertheless, there remain a number of problems for this type-based interpretation of DPT. Now, consider three more criticisms:

Criticism 4: Types do not distinguish mental processes. The restructuring of DPT based on processing types was largely intended to solve the cross-cutting problem by using the continuity of mental processes to its advantage. This works insofar as it sidesteps the issue of having to carefully demarcate separate systems; but it essentially pushes the problem back a level and does not provide a solution to the ambiguity surrounding mental processes (Keren 2013). While proponents like Evans and Stanovich may argue that working memory is a sufficient criterion to distinguish the autonomous set of systems from non-autonomous ones, this does little to improve understanding of the putative menu of processes which make up Type 2 functionings. This "stripped down" version of DPT makes the overall framework less precise, which makes one wonder whether it is not simply a theory about working memory instead of a theory about reasoning and decision-making (Keren 2013; Mugg 2016). I return to this issue in section 4.2.

Criticism 5: The criterion of "rule-based" reasoning is ambiguous. Both system-based and type-based interpretations of DPT appeal to processes that are "rule-based". As argued by Kruglanski and Gigerenzer (2011), there seems to be much equivocation in the use of the term "rule-based" as a criterion to distinguish types of processes. On the one hand, "rule-based" could refer to the conscious effort of an individual to adhere to rules (e.g., rules of normative conduct, rules of a game, rules of arithmetic); but, on the other hand, "rule-based" could refer to unconscious "computational" processes that aid in or underwrite cognitive tasks. For some, namely those interested explicitly in the psychology of deductive reasoning, this equivocation may not be much a problem, the property "rule-based" refers typically to higher-order capacities to reason abstractly and perform mental simulations. However, for those interested in the mental processes that support the learning of implicit skills and other preconditions for reasoning and decision-making, this can get confusing very quickly, as it's not obvious whether the criterion refers to a conscious ability of the individual to reason analytically, or whether it refers to the ability to model some aspect of cognition according to rules.[8] Evans and Stanovich

[8] The discrepancy here about what it means to describe mental processes as 'rule-based' runs parallel to debates in philosophy of cognitive science concerning the meaning and interpretation of "computation" in computational theories of cognition (Van Gelder 1995; Thompson 2007; Miłkowski 2013a, b; Piccinini and Bahar 2013; Piccinini 2015). The bone of contention is whether, or on what grounds, it makes sense to say that the mind computes information, i.e. at the level of representations or at the level of neurophysiology (or in between). Skeptics of computational theories of mind reject claims that human thinking is computational because there is, as of yet, no evidence that anything akin to symbol-manipulation happens when thought is produced (Hutto and Myin 2013, 2017; Hutto et al. 2018; cf. Colombo 2014).

(2013a, b) are not convinced this is a major issue, but criticism (6) indicates further why it may turn out to be a worthy criticism.

Criticism 6: What does algorithmic processing refer to? The idea of algorithmic processing was introduced to alleviate confusions about where and how Type 2 processes are initiated; Stanovich (2009, 2011) has argued that this innovation has been very useful for explaining how implicit skills are developed and for accounting for individual differences in cognitive ability and intelligence. This very well may be the case. But our problem is that it's anything but clear how algorithmic processes are realized, and how they differ—at the neuroanatomical level—from other Type 2 processes, if they do at all. Stanovich uses primarily functional terminology to portray the rule-based nature of algorithmic processes—but this doesn't alleviate the problem of mental processes cross-cutting each other, nor does it clarify what it means to describe some mental processes as rule-based. We are told that even though individuals are not conscious of algorithmic processes, they are still considered Type 2 processes because they depend on working memory and are representational in nature.

In sum, these additional criticisms suggest that type-based interpretations DPT may obscure rather than clarify the idea that human reasoning and decision making is *inherently* dualistic. Part of the reason for this is that many of the same reasoning processes can be also described by a 'one-system' model (Osman 2004; Kruglanski and Gigerenzer 2011) or, as I mentioned above, models with many systems and stages (Stanovich 2011; Varga and Hamburger 2014; Pennycook et al. 2015; Swan et al. 2018). This can be seen as a further justification for the claim that rational decision-making cannot be reduced to the operations of single system, or in this instance, a single type: Type 2 processes, like System 2 processes, do not guarantee rational decision-making, and likewise, Type 1 processes, like System 1 processes, do not necessarily produce irrational actions.

4 Two Styles of Dualistic Decision Modeling

In this section, I review two common styles of dualistic decision modeling in economics. The first style sees DPT identifying neural structures in the brain; the second style sees DPT identifying functional characteristics of mental processes. Regarding the former, I argue that because the standard view of DPT is not bound to any particular neuroanatomical interpretation, dualistic decision models which utilize neuroscience evidence may exaggerate or distort the roles that specialized mechanisms and/or subsystems play in the evaluation of prospects and execution of decisions. This puts into perspective the relevance of criticisms (1–3) above for neuroeconomic research. Regarding the second style, I argue that when DPT is construed as a functional (non-reductive) framework, it cannot explain how choice emerges from the interaction of dual processes and/or dual systems without a supplemental story. This leaves behavioral economists with two options: (i) bite the bullet and leave the black box closed; (ii) use formal constructs from economics to work out the details of how dual processes and/or systems interact. The problem with the latter option is that the explanatory value of dualistic decision models becomes hostage to the explanatory value of DPT in

general. This puts into perspective the relevance of criticisms (4–6) above for behavioral economic research.

4.1 Dualistic Decision Models in Neuroeconomics

Camerer et al. (2005) (CLP) have championed the use of neuroscience to improve economic theory. They emphasize that standard economic theory is inadequate because it is unable to account for decision anomalies that result from "automatic" and "emotional" processing which governs an extensive portion of human behavior. "Human behavior," they argue, "requires a fluid interaction between controlled and automatic process, and between cognitive and affective systems. However, many behaviors that emerge from this interplay are routinely and falsely interpreted as being the product of cognitive deliberation alone" (CLP, 2005, p. 11). Not surprisingly, one finds many references to core texts from the DPT literature in support of CLP's "two-dimensional" neuroeconomic framework (e.g., Shiffrin and Schneider 1977; Sloman 1996; Kirkpatrick and Epstein 1992; Lieberman et al. 2002; Gollwitzer et al. 2004; Kahneman and Frederick 2002); yet, CLP are also adamant that the contrastive functions of automatic and controlled processing are indicative of separate emotional and cognitive systems in the brain.

Unlike dualistic decision models in behavioral economics which merely pay lip-service to neural evidence (see section 4.2), CLP believe that neuroscience can illuminate how individuals form their preferences and can improve how economists study rational choice.[9] In searching for a neural basis for rational choice, CLP have come to emphasize the importance of specialized mechanisms, such as the "executive control" and "conflict monitoring" mechanisms, for understanding what differentiates controlled from automatic processing. These mechanisms are taken to be responsible for initiating "override" functions which suppress automatic impulses to act or consume. For this reason, CLP have come to associate executive control and conflict monitoring mechanisms with successful impulse control and delayed gratification in intertemporal choice contexts:

> Controlled processes occur mainly in the front (orbital and prefrontal) parts of the brain. The prefrontal cortex (pFC) is sometimes called the "executive" region, because it draws inputs from almost all other regions, integrates them to form near and long-term goals, and plans actions that take these goals into account. (Camerer et al. 2005, p. 17)[10]

[9] CLP understand the notion of "rational behavior" as indicating a particular sort or kind of processing, namely as deliberate and under conscious control. This is different from understanding "rational behavior" as behavior being consistent with a set of axioms.

[10] Similarly, McClure et al. claim that: "We have observed that choices between lesser immediate and greater delayed rewards elicit activity in distinct neural systems that appear to favor different choice outcomes. That is, intertemporal choice under these conditions elicits decisional conflict. A growing body of evidence suggests that a dorsocaudal region of the ACC [dorsal anterior cingulate Cortex] responds to conflicts in processing… This is consistent with findings from the current study in which we observed activity in a similar area of the ACC that was greater for decisions involving choices between immediate and delayed rewards than for choices between only delayed rewards. Such findings have been taken as evidence for a conflict-monitoring function of ACC, which serves to detect conditions requiring the recruitment of cognitive control mechanisms subserved by prefrontal cortex and associated structures" (2007, pp. 5803).

For CLP, the mechanisms in question not only provide a clearer picture of how the brain works, but they serve as a vehicle to track valuation procedures among separate neural systems. This same view has been endorsed by other neuroeconomists, including McClure et al. (2004, 2007) and Brocas and Carrillo (2008, 2014).

However, it's easy to lose track of how the executive control and conflict monitoring mechanisms relate to the DPT framework. For neuroeconomists like CLP and McClure et al, the Cognitive System is analogous to System 2; hence, if executive control is what prevents automatic and impulsive processing from running wild, then the mechanisms responsible for executive control should be contained by, that is operate within, the Cognitive System. But we may then wonder how these mechanisms resolve conflict between separate cognitive and emotional neural systems and how they relate to deliberation and rational choice. To take another example, Brocas and Carrillo (2014) argue that the "central executive system" coordinates the tasks of separate sub-systems by governing the flow of information between regions of the brain (cf. D'Esposito et al. 1995; Szameitat et al. 2002). Like CLP and McClure et al., Brocas & Carrillo also regard the central executive system as crucial to understanding rational choice formation. Yet, the central executive system for Brocas & Carrillo is not under the individual's control—it operates autonomously and is not accessible to introspection; for this reason, it's hard to know how conflict monitoring and executive control at the neural level causally relate to controlled mental processing that we typically associate with rational choice. It leads one to wonder whether the cognitive-emotional duality endorsed by CLP and other neuroeconomists is really necessary to their purposes—after all, it is the conflict monitoring and executive control mechanisms which are doing the explanatory work, not the individual who must contemplate having immediate or delayed rewards. These examples illustrate an important conceptual gap between how DPT is conceived as a descriptive framework, and how neuroscientific evidence fits into that framework.

Now, just because the standard view of DPT is not committed to a particular neural interpretation doesn't preclude one from speculating about the neural foundations of certain functions, such as cognitive control. But if one invokes specialized mechanisms or sub-systems then we ought to have some idea how these support DPT. CLP's and McClure et al's two-system models are ambiguous with regard to what *really* determines individual's preferences— the cognitive and emotional systems? The controlled and automatic processes which comprise those systems? Or the executive control and conflict monitoring sub-systems which govern the controlled and automatic processes? There is clearly some sort of hierarchy here; but how it relates to and informs the economic analysis of rational choice is left unspecified. Rather, CLP say things like this:

> Neuroscience is shot through with familiar economic language—delegation, division of labor, constraint, coordination, executive function—but these concepts are not formalized in neuroscience as they are in economics. There is no overall theory of how the brain allocates resources that are essentially fixed (e.g., blood flow and attention). An "economic model of the brain" could help here. Simple economic concepts, like mechanisms for rationing under scarcity, and general versus partial equilibrium responses to shocks, could help neuroscientists understand how the entire brain interacts. (Camerer et al. 2005, p. 56)

There is surely merit to this claim as neuroeconomics is developing into an independent sub-field of behavioral economics. But the question that arises is, if the logic of economic theory is suited to analyze decisions as they are realized in the brain, as CLP argue, then where—at what point in the causal decision process—should economic models focus? At the penultimate moment of execution? Or across different decision nodes? This question is important not just for understanding how dualistic decision models envision the use of neuroscientific evidence, but also for understanding what rational choice consists in for models like CLP's.

It is useful to introduce a distinction here between two styles of neuroeconomics. The approach of CLP and McClure et al., which uses neuroscientific methods to elicit and characterize individuals' preferences, has been dubbed "behavioral economics in the scanner" (BES) (Harrison and Ross 2010). This approach can be contrasted with what has alternatively been called "economics of neural activity" (ENA) (Vromen 2011) which uses econometric methods to study the neurobiology of the brain. It is the former style of neuroeconomics that we are here concerned: Unlike ENA, proponents of BES rely on dualistic properties to depict where and how rational choice is executed (when it is executed properly). According to proponents of ENA, the idea of partitioning the brain into spatially distinct valuation systems, as CLP and McClure et al. do, is misguided—there is really one single valuation system (cf. Parker and Newsome 1998; Thompson and Schall 1999; Glimcher 2003).

Vromen (2011) suggests that the disagreement between proponents of BES and ENA over how many valuation systems there are in the brain can be reconciled by instead thinking in terms of "*upstream* and *downstream* phases in the total causal chain of decision-making in the brain" (2011, p. 278, my emphasis). Hence, it may after all be the case that there are multiple regions and circuits responsible for "upstream" valuation which fit CLP and McClure et al's dualistic picture of the decision-maker; but along the way, these valuations converge to a single phase or node, which requires an alternative picture of the execution processes, and hence, an alternative model of the mind / brain.

The message to take away here is not that economic models cannot interface with neuroscience to produce more realistic models of individual decision-making; but doing so via DPT runs the risk of distorting (by overemphasizing) the role of certain mechanisms in that process. This speaks to the significance of criticisms 2 and 3 above, which suggests that not only is the story of the interaction of systems descriptively ambiguous, but also that this story is underdetermined by contemporary neuroscientific evidence. For those who practice BES-style neuroeconomics, the mechanisms responsible for executive control and valuation (which comprise the cognitive and emotional systems and which produce controlled and automatic processes) are only part of the story. *If* CLP and McClure et al. ultimately agree with proponents of ENA that the final phases of decision-making are integrated into a single causal stream, then DPT is useful *only* insofar as it potentially represents upstream stages of decision-making prior to integration. After integration, only one system is in operation.

One might then wonder why neuroeconomists need DPT at all given that its representation of decision-making processes is (at best) incomplete and (at worst) amenable to distortion. According to CLP and McClure et al., it is only through studying the contrastive features of controlled and automatic processing in the brain that behavioral economists will be able to refine and reform standard economic theory.

Perhaps there is something to this. But it is far from evident that DPT is best or even a reliable framework for depicting such processes in the brain.

4.2 Dualistic Decision Models in Behavioral Economics

The section above summarizes a limiting set of cases: it indicates that DPT may be descriptively inadequate to represent and analyze decision processes in the brain. But the case could be made that if these inadequacies constitute a problem at all, it is only a problem for neuroeconomic applications of DPT, not for behavioral economics in general. Rather, it is much more likely that behavioral economists view DPT as a functional (non-reductive) psychological framework, one which provides *approximate* descriptions of decision procedures based on the standard menu of mental processes (or something like that). On this view, the predictive and/or explanatory power of dualistic decision models is independent of whatever neural structures are involved in decision-making and thus is unaffected by such descriptive problems. This functional interpretation would be closer in-step with the putative aims of behavioral economics according to, e.g., Angner and Loewenstein (2012) and Angner (2019). Yet, I will argue that this functional interpretation nevertheless generates more, not less, confusion about the role that DPT plays with regard to behavioral economists scientific ambitions. There are two aspects to this.

First, how dualistic decision models in behavioral economics understand the interaction of dual processes and/or systems is rarely if ever explicitly formulated. Recall that, in addition to the lack of neural evidence for demarcating dual processes and systems in the brain (see criticisms 2 and 3), criticism 1 indicates that mental processes operate on a continuum and are not functionally discrete—which is to say different mental processes may crosscut each other or operate jointly to serve different functions. While proponents of DPT (such as Evans and Stanovich 2013a) see this as an effect of the generalization of the standard menu of mental processes, it's important that researchers understand that the crosscutting problem is an inherent challenge for any functional interpretation of DPT. For this reason, the story of how dual processes and/or systems interact is left in the black box.

While some behavioral economists seem to be willing to bite the bullet and forego providing any explanation of system interaction, others have attempted to overcome this issue by interpreting mental processes formally, using the tools of economic analysis to flesh out the details. For example, it has become a common tactic among behavioral economists to use principal-agent models and limited-information games to represent the internal dynamics of decision-making in the context of intrapersonal and intertemporal choice—see, e.g., Bénabou and Tirole (2002), Bernheim and Rangel (2004), Benhabib and Bisin (2005), Loewenstein and O'Donoghue (2005), Fudenberg and Levine (2006). As such, what determines whether one system or type of processes overrides and intervenes on another is contingent upon the bargaining or constraining power of that system or type of processes construed as an economic agent. Familiar psychological concepts such as *willpower* and *cognitive control*, which are often associated with the capacity for rational self-control, are thus derived from the mathematics of the model. The details of *how* DPT's functional characteristics *may* map onto these formal constructs has been discussed by Grayot (2019), so I will not repeat

them here. Yet, the central raised by Grayot, which is applicable here, is whether these formal constructs possibly misconstrue how dual systems and/or dual types of mental processes produce choice at all. To understand how this relates back to the first three criticisms, consider again the parallel-competitive and default-interventionist interpretations of system interaction.

Loewenstein and O'Donoghue (2005) propose a model of system interaction which is quite similar to the default-interventionist model. Their model presumes that the Affective System is active by default, whereas the Deliberative System requires energy resources to play the part of monitor and intervener. Now, why they adopt this model and not something closer to the parallel-competitive model is unknown, since there is not much background discussion about it. One reason to forego the parallel-competitive model might be that the default-interventionist model presumes the rational authority of the Deliberative System (the Deliberative System has a kind of agency which enables it to propose a menu of choices to constrain the Affective System); the parallel-competitive model of system interaction does not lend itself to such an interpretation. In fact, Loewenstein and O'Donoghue (2005) state that their model relegates the impulses and motivations of the Affective System to an exogenous variable, which is fixed independently of the valuations of the Deliberative System (2005, p. 6). One gets the impression the default-interventionist model is popular among behavioral economists not because it is the most realistic (supported by neuroscientific evidence), but because it is the most flexible with regard to its application to different choice contexts. Here's why:

The criteria by which DPT distinguishes the functional characteristics of mental processes are few and open-ended. While there are endless number of interpretations of DPT in the cognitive sciences and psychology, the case has been made that what ultimately demarcates slow, controlled, and reflective mental processes from fast, reactive, and automatic ones rests on just basic two criteria: *autonomy* of operation and the requirement of *working memory* (Evans and Stanovich 2013b; cf. Thompson 2013). On this interpretation, it's easy to see how DPT can be leveraged to characterize any number of decision anomalies.

But this gives rise to a more fundamental question, namely, if autonomy and working memory are all that is required to justify demarcating and ascribing mental processes to some choice phenomenon, what does it take to possibly refute DPT? Consider the following passage by Pennycook:

> The observation that the distinction between intuition and reflection is irrefutable is foundational because it means that dual-process models should not be concerned with justifying this claim. That is, dual-process models must take this distinction as a given and build from there. If we know with a reasonable degree of certainty that the mind has this capacity for two different types of processes (autonomous and non-autonomous), where do we go from there? […] Thus, the mere distinction between intuition and reflection based on autonomy is sufficient for the claim that dual-process theory is irrefutable, but not sufficient for the claim that the theory is worth anyone's time. (2017, p. 8)

Although Pennycook does not regard this foundational aspect of DPT to be inherently problematic,[11] this should raise at least some alarms with regard to the scientific ambitions of behavioral economists. One need not be card-carrying Popperian to see that irrefutability may run contrary to the aims of behavioral economics, which are "to improve the realism and psychological assumptions underlying economic theory" (Camerer 1999). If DPT is thus construed as a functional framework, its theoretic structure allows for virtually unlimited interpretation. I return to this issue in the next section.

Of course, one could retort that DPT is better construed as a *meta-theory* than a first-order theory. As Evans & Stanovich continue to argue in defense of DPT, "such frameworks [meta-theories] cannot be falsified by the failure of any specific instantiation or experimental finding. Only specific models tailored to the tasks can be refuted in that way…" (2013b, p. 263). According to Evans & Stanovich, this is precisely why the System 1 /System 2 distinction is misleading—it gives the impression that the various dichotomies underlying DPT are strict and consistent across token theories—which we now know not to be the case. Hence, they abandon the system-based interpretation of DPT and commit to (what they believe is) a single, more coherent, dichotomy between autonomous processes and those which require working memory—this, recall, is the Type 1/Type 2 distinction, which they endorse as the least problematic interpretation of DPT (cf. Evans 2017).

5 DPT and the Myth of the Inner Rational Agent

Where does this leave us? It would seem that, due to the descriptive inadequacies of DPT, both styles of dualistic decision modeling have considerable limitations. One may then ask: what justifies taking a dualistic perspective about decision modeling at all? In this section, I consider an alternative reason for DPT's popularity, one which trades on its fundamental irrefutability. I speculate that DPT appeals to behavioral economists because it satisfies modeling needs that are normative in origin: for economists, DPT is not a theory about how decisions are made, but a theory about why rational agents tend to deviate from the core tenets of microeconomic rationality. On this interpretation, System 2 (and its various analogs) corresponds to an *inner rational agent*, one that would otherwise abide by the norms of expected utility theory were it not for the failures of System 1.[12] Although I take this to be the most plausible reason for DPT's popularity in economics, I will argue that even this interpretation is not entirely justified. In particular, I argue that neither system-based nor type-

[11] He states, the "true test of a good theory is whether it can be applied successfully to problems and generate hypotheses" (Pennycook 2017, p. 8).

[12] The arguments developed in this section run parallel to those presented in Sahlin et al. (2010) who argue that decision theorists have often treated System 2 as an "approximation" of what Edwards (1954) referred to as "Normative Man". I am highly sympathetic to Sahlin, Wallin, and Persson's claims that DPT lacks a "firm conceptual framework" and is inadequate to the purposes of decision theory, especially in the tradition of Prospect theory. However, as shown in section 2, the links between DPT and economic decision research extend beyond decision theory. Thus, to avoid conflating my own arguments with theirs, I have adopted the term "inner rational agent" from Infante et al. (2016) who likewise observe that behavioral economists have come adopt a dualistic perspective of human agents, one in which the rational economic agent is trapped inside an "outer psychological shell".

based interpretations of DPT establish a clear basis for demarcating rational from irrational decision processes. This puts into perspective and exemplifies the second set of criticisms raised in section 2, viz. criticisms 4–6.

5.1 System 2 / Type 2 Processes aren't Necessarily Rational

For ordinary persons, what counts as rational action is a matter of degree and often context-dependent. The same cannot be said of economic agents—rationality is judged according to whether choice-behavior is consistent with expected utility theory. For this reason, the contrastive features of DPT appear to be tailor-made for economists: having the ability to distinguish between mental processes that generate reasoning errors and mental processes that don't is a critical tool for the analysis of rational choice. As already indicated, many consider System 2 (or some analog of it) to be the "rational system": this is not just because it supports higher cognitive functionings, like hypothetical and counter-factual thinking, but also because it is associated with the detection and inhibition of biased judgments and impulsive behaviors. But is it safe to presume that System 2 always produces rational outcomes? The answer is no; but to understand why, we need to clarify things.

Firstly, when proponents of DPT refer to the rational capacities of System 2, this is based on a standard of rationality that is rooted in the norms of deductive logic and probability theory (Evans and Over 1996; Gigerenzer 1996; Stanovich 1999; Stein 1996). Although this normative standard has been a point of much contention in the philosophy and psychology of human reasoning (cf. Gigerenzer and Goldstein 1996; Samuels et al. 2012; Samuels, & Stich 2004; Over 2004), we can set it aside for now. Secondly, System 2 is often identified with *critical thinking*. Critical thinking, according to the American Philosophical Association, is defined as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual consideration upon which that judgment is based" (Facione 1990). But what does it mean to identify System 2 with successful critical thinking? It could mean that (i) System 2 is a necessary condition (prerequisite) for critical thinking; or it could mean that (ii) System 2 is sufficient for critical thinking, which is to say, that all instances of System 2 processing are instances of critical thinking (Bonnefon 2018). It's not difficult to imagine why the latter interpretation isn't realistic. Not only is System 2 operative during all ordinary activities that don't meet requirements of critical thinking (e.g., reading a book engages many System 2 processes, like mental simulation and hypothetical thinking, but this does not count as an instance of critical thinking); more importantly, there are instances where higher cognitive functions, like reflection and deliberation, cause people to make mistakes that may not have otherwise occurred. Bonnefon (2018) gives two examples of errors that result solely from System 2 processing: one is related to false justification or what he calls "pseudo-rational" answers. Often people *follow* their initial impulses and seek to justify them through clever rationalizations. This would support classic studies by Nisbett and Ross (1980) which reveals that individuals may give

false verbal reports to justify actions they had no control over—often they do this without realizing the report is false. Another example of System 2 failure is the result of over-thinking, in which a person may mix-up or confuse relatively simple information by deliberating on it.[13] While these examples of System 2 errors are not as systematic as those commonly attributed to System 1 by proponents of heuristics and biases research, they illustrate an important point, which is that System 2 processing does not guarantee critical thinking and, moreover, does not prohibit reasoning errors from occurring. For this reason, it would be mistaken to presume, as many dualistic decision models tacitly do, that System 2 is rational by default, or rather, that if System 2 is rational then engaging system 2 precludes reasoning errors.

How does this relate to DPT and the concept of the inner rational agent? An important aspect of many system-based interpretations of DPT is that individuals make reasoning errors when System 2 does not have the resources to monitor or inhibit System 1 functions. A successful completion of System 2 thinking is said to pass three stages, *conflict detection*, *sustained inhibition*, *explicit resolution* (De Neys and Bonnefon 2013; Pennycook et al. 2015; Stanovich and West 2009). Yet, if one thinks that rationality is defined according to the norms of logic and probability theory, then it reveals that System 2 does not guarantee rational action—in fact, because rationality is based on coarser standards than critical thinking, it's likely that violations of rationality by System 2 are more common than instances of non-critical thinking—one can very easily make calculation errors while thinking critically about a decision. Of course, this only addresses the strictest associations of System 2 with rational choice.

Let's now consider the former disjunct, that System 1 is the primary cause of reasoning errors. While the above argument implies that System 1 is not the sole cause of reasoning errors (viz. because System 2 is sometimes involved), proponents of "ecological rationality" (Gigerenzer et al. 1999; Gigerenzer 2004, 2007, 2008; Gigerenzer and Brighton 2009) argue that it is inappropriate to presume that rationality is constituted by the norms of deductive logic and probability theory. Many useful heuristics are generated by mental processes associated with System 1, not all of which lead to reasoning errors. Even if these processes are evolutionarily old and not conducive to modern choice contexts, automatic and implicit processing is necessary for many higher-reasoning tasks. Yet, the reason why decision researchers and economists treat *these* processes as inherently irrational is because they help to predict a small range of decision phenomena that are relevant for some economic purposes. Hence, supporters of ecological rationality maintain that DPT presupposes its normative assumptions: if the reason for adopting System 1 and System 2 is that it adheres the normative standards of deductive logic and probability theory, this begs the question (Gigerenzer and Brighton 2009; Gigerenzer and Sturm 2012).

But perhaps this is only a problem for the system-based interpretation of DPT, which we've now seen to have difficulties demarcating separate mental processes. What if

[13] Support for both kinds of System 2 errors are discussed in Mercier and Sperber (2011) who argue that one of the functions of System 2 is the production of arguments. In this sense, an argument is a complex representation of propositions and a conclusion which is derivable from those propositions. It's possible for a well-functioning System 2 to incur reasoning errors as a result of trying to convince others of an argument rather than trying to arrive at true beliefs. This explains problematic rationalizations and other forms of confirmation bias (see Mercier and Sperber 2011, pg. 63–66, for discussion).

economists were to adopt a subtler, less metaphysically loaded framework for the analysis of rational choice? For instance, would the type-based, as opposed to system-based, interpretation of DPT help alleviate these issues? The answer is probably not.

Firstly, proponents of the type-based interpretation of DPT do not agree on the source of reasoning errors. As critics have argued, this may have something to do with the fact that the Type 1/Type 2 distinction does not solve the cross-cutting issue (as described by criticism 1); instead, it compounds the issue by generalizing the functional characteristics of Type 1 and Type 2 mental processes. For this reason, criticism 4 extends the cross-cutting issue to include type-based interpretations of DPT, rendering them incapable of attributing reasoning errors to specific mental processes.

Secondly, even more sophisticated versions of the type-based interpretation of DPT, such as Stanovich's tri-process model (2009; 2011) seem to be unable to attribute reasoning errors to specific mental processing types. Recall that, for Stanovich, Type 2 processes are comprised of two stages, an algorithmic stage and a reflective stage. Unlike the reflective stage, the algorithmic stage is not accessible to introspection; it operates below the threshold of conscious awareness and is primarily responsible for initiating override procedures that persons then experience (at the reflective stage) as exerting cognitive control or willpower or whatever it is that enables them to stave off impulsive action and perform rational actions. It thus may be tempting to think that the algorithmic stage is partly responsible for some reasoning errors, and further, that the reflective stage is ultimately responsible for rational action. The reason that this tri-process model can't help economists establish which mental processing types are responsible for reasoning errors is because it doesn't commit to a clean distinction between the algorithmic and reflective processing. As criticisms 5 and 6 indicate, though there are plenty of candidate structures at the neuroanatomical which *could* support algorithmic stage operations, Stanovich's portrayal of the algorithmic stage eschews most neural interpretations.[14] Given that he understands algorithmic processing as a set of functional, rule-bound procedures that facilitate reflective processing, there is no theoretical reason why the reflective stage of Type 2 couldn't also be involved in reasoning errors. When individuals (unknowingly) justify mistakes or attempt to assimilate impulsive behaviors with pseudo-rational answers, this must pass through the reflective stage. This serves as a further reason for not giving Type 2 processes full rational authority.

For the purposes of rational choice analysis, it thus seems that neither system-based nor type-based interpretations of DPT justify a belief in the concept of an inner rational agent. Even if behavioral economists concede that the interactions between System 1 and System 2 (or Type 1 and Type 2) are only an approximation of the internal dynamics of decision making, it would require significant deviations from DPT to

[14] Stanovich (2011) uses psychometric data to make a compelling case for the role of the algorithmic stage in Type 2 processing (even if it's not clear what underwrite those processes). His argument rests on claims that individual differences in cognitive ability are indications that something, prior to reflective processing, has override control of Type 1 processes. The reason why these processes simply aren't relegated to Type 1 is because they are not intrinsically autonomous; rather, they are learned and internalized with practice. Reading comprehension and arithmetical skills are examples of learned skills. The case has been made that such data relies on question-begging assumptions and driven by confirmation bias (Polonioli 2014).

continue to promote dualistic decision models which presume that somewhere, within the individual, is a rational agent.

6 Concluding Remarks: On Scientific Ambitions and Normative Commitments

To recap, I've argued that behavioral economists are faced with a dilemma: either they stick to their purported ambitions to give a realistic description of human decision-making and give up the dual process narrative, or they revise and restate their scientific ambitions. Section 2 surveyed both the explicit and implicit influences of DPT upon economics and posited that there are two general styles of dualistic decision modeling, behavioral economic and neuroeconomic. In section 3, I provided an overview of the current status of DPT from the perspective of cognitive science and philosophical psychology; I elaborated six criticisms of DPT, three of which pertain to system-based interpretations (upon which the standard menu of mental processes is commonly defined), and three more criticisms which pertain to recent type-based modifications of DPT. The take-away message of section 3 is that DPT is not as descriptively accurate as it is often portrayed to be, and therefore, not a reliable framework for depicting mental processing: it posits structures that it cannot articulate and processes that it cannot track. Section 4 considered whether the use of DPT generates tensions for either of the two styles of dualistic modeling and discusses their possible limitations. In Section 5.1 I argued that, at best, DPT provides a narrative upon which behavioral economists can base their models—the idea that within each person is an inner rational agent allows economists to interpret deviations from expected utility theory as the outcome of irrational or non-rational processes. But even this narrative role would require further justification given that System 2 / Type 2 processing can incur reasoning errors (and in some contexts, System 1 / Type 1 processing may be sufficient for rational action). In sum, not only is DPT subject to intense scrutiny as a general theory of mental processing, but the versions of DPT that many economists subscribe to is a not a faithful representation of either system-based or type-based interpretations of it. They are often caricatured versions of DPT.

However, one can envision a few reasons why behavioral economists and neuroeconomists may not be convinced by the arguments above, and further, may not wish to give up on the dual process narrative. For instance, it could be argued that despite its theoretical challenges, DPT is nevertheless a plausible (meta-)theory, and this is all that behavioral economic research demands of its psychological foundations. This line of reasoning is to be expected. For instance, Angner & Loewenstein claim that:

> These days, as it is typically employed, "behavioral economics" refers to the attempt to increase the explanatory and predictive power of economic theory by providing it with more psychologically plausible foundations. (2012, p. 1)
>
> …many behavioral economists believe that social and behavioral science should aspire to reveal the actual causes of behavior… On this view, the psychological plausibility of underlying assumptions, and the accuracy of predictions matter for the assessment of a theory. They matter because psychological plausibility and

> predictive accuracy are seen as indications that the theory "has pinned down the right causes," not because they matter in and of themselves. (2012, p. 36-7)

In fact, behavioral economists may not be willing to give up on DPT precisely because it's way of partitioning the mind/brain is naturally attuned to the needs of economic analysis. The idea that System 2 (Type 2) would otherwise make rational decisions were it not for System 1 (Type 1) allows economists to retain the neoclassical ideal of rationality while accounting for deviations from this ideal. In other words, it establishes a point of reference from which less-than-rational behavior can be measured and modeled. This point has been forcefully argued by Berg and Gigerenzer (2010) who claim that behavioral economics is just neoclassical economics in "disguise". While it is beyond the scope of this paper to discuss whether the normative standards of neoclassical rationality are psychologically justified, and whether the caricature version of DPT that dualistic economic models often subscribe to corresponds to normative standards of rationality held by proponents of DPT in psychology and cognitive science, it is worth considering what a refusal to give up DPT says about the scientific ambitions of behavioral economists (assuming they are not swayed by theoretical and empirical criticisms raised above).

Berg & Gigerenzer question this scientific ambition in the following inquiry:

> Insofar as the goal of replacing [the] idealized assumptions [of neoclassical economics] with more realistic ones accurately summarizes the behavioral economics program, we can attempt to evaluate its success by assessing the extent to which empirical realism has been achieved. Measures of empirical realism naturally focus on the correspondence between models on the one hand, and the real-world phenomena they seek to illuminate on the other… given its claims of improved realism, one is entitled to ask how much psychological realism has been brought into economics by behavioral economists (Berg & Gigerenzer 2012–2)

Berg & Gigerenzer assert that behavioral economics does not achieve empirical realism on multiple fronts.[15] But, what stands out between Angner and Loewenstein (2012) and Berg and Gigerenzer's (2010) appraisal of the scientific aims of behavioral economics is the difference between commitments to psychological *plausibility* and psychological *reality*. Apologists of the dual process narrative may be inclined to agree with Angner & Loewenstein that because DPT is plausible and because it affords better predictions and explanations of decision phenomena than neoclassical models, there is nothing scandalous in choosing to overlook its known deficiencies. Afterall, if psychological plausibility and predictive accuracy are indications that the theory "has pinned down the right causes" (cf. Camerer et al. 2003, 4), then this seems to be justification enough to stick with DPT. Apologists may dismiss Berg & Gigerenzer's criticisms from empirical realism as exceeding

[15] They characterize this by way of three methodological limitations: (i) "restrictions on what counts as an interesting question," (ii) "timidity with respect to challenging neoclassical definitions of normative rationality," and (iii) "confusion about fit versus prediction in evaluating a model's ability to explain data" (Berg and Gigerenzer 2010, p. 3).

the call of economic duty. The psychological plausibility of DPT permits one to model decisions as the outcome of dual processes.

The problem with this line of reasoning is that it rests on a subtle equivocation—namely, that the *right* causes correspond with the *actual* causes. While behavioral and social science may aspire to identify the actual causes of decision phenomena, as Angner & Loewenstein claim, both economists and psychologists enjoy a certain latitude in deciding what are the right causes according to their favorite theories. I've argued throughout this paper that DPT is not descriptively accurate with regard to the actual causes of decision phenomena; and I suggested above in section 5.1 that DPT's staying power in behavioral economics is due to the fact that it supposedly provides psychological foundations for rational choice by treating System 2 (Type 2) as an inner rational agent. But if these foundations are based on a caricature of DPT, one which oversimplifies the role and function of disparate mental processes in the production of rational (and irrational) behavior, then the right causes have effectively come apart from the actual causes.

If the right causes don't reflect the actual causes, then we should like to know how DPT increases the explanatory and predictive power of economic theory? If neoclassical economics is taken as the descriptive benchmark, then of course dualistic decision models constitute increased explanatory and predictive power. But the challenge to answering this question is that we often don't know what is the benchmark against which a model successfully predicts or explains phenomena. This paper has provided plenty of reasons to think that DPT is descriptively inaccurate regarding the actual causes of decision-making processes—I therefore leave it an open question whether dualistic economic models serve alternative scientific purposes.[16]

What about the increased predictive power of dualistic economic models? A recurring criticism in Berg and Gigerenzer (2010) (see also Gigerenzer 2015) against behavioral economics is the failure of researchers to provide genuine out-of-sample predictions. They argue:

> Given that many behavioral economics models feature more free parameters than the neoclassical models they seek to improve upon, an adequate empirical test requires more than a high degree of within-sample fit… Arguing in favor of new, highly parameterized models by pointing to what amounts to a higher R-squared (sometimes even only slightly higher) is, however, a widely practiced rhetorical form in behavioral economics. (2010, p. 15)

Berg & Gigerenzer proceed to give examples of this rhetorical practice in different domains of economics, including judgment and decision-making in the Heuristics and Biases tradition as well as intertemporal choice and hyperbolic discounting. While not all of their target models are dualistic, there is an affinity between their charges against behavioral economists who overstate the predictive power of their models and the criticisms raised here against DPT in social and cognitive psychology. The take away message here is not that dualistic decision models in behavioral economics couldn't make novel predictions or couldn't provide useful explanatory frameworks for organizing behavioral data; but that, insofar as we don't have clearly articulated

[16] Though, Heilmann (2016) raises doubts about the unificatory potential of DPT for economic purposes.

success conditions, the known theoretical and empirical deficiencies of DPT call into question *how* exactly dualistic models predict and explain decision phenomena.

Finally, one might respond that my criticisms of dualistic decision models and my argument that economists are confronted with a dilemma over-generalizes behavioral economic practices—that, at best, the deficiencies of DPT have only limited relevance to the disciplines on the whole. I am sympathetic to this response. Indeed, it was not the aim of this paper to make grand assertions or presumptions about the goals, methods, and/or theoretical commitments of behavioral economists or neuroeconomists *in general*. As indicated in the first two sections, this is because behavioral economics is a heterogenous discipline; much of what is now recognized to be bona fide behavioral economic research has its roots in experimental psychology as well as cognitive (neuro)science, and the increase in interdisciplinarity has led to fuzzier borders separating economic from non-economic concepts and evidence. It would, in fact, make little sense to apply any broad generalizations—critical or otherwise—to behavioral economics.

Nevertheless, the dilemma I've posed in this paper indicates that if behavioral economists and neuroeconomists are not swayed by the growing criticisms against DPT, then the burden of justification is upon them to clarify what are their scientific aims in utilizing DPT. Economists are of course permitted to pick and choose whichever theoretical frameworks best suit their modeling needs; but the scientific community is also permitted to inquire what is the scientific value of their models, especially when there are known deficiencies in their theoretical assumptions.

References

Ainslie, G. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person*. Cambridge University Press.

Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press.

Ainslie, G. 2005. Précis of breakdown of will. *Behavioral and Brain Sciences* 28 (5): 635–650.

Ainslie, G., and N. Haslam. 1992. Self-control. In *Choice over time*, ed. G. Loewenstein and J. Elster, 177–209. New York: Russell Sage Foundation.

Angner, E. 2019. We're all behavioral economists now. *Journal of Economic Methodology* 1–13.

Bénabou, R., and J. Tirole. 2002. Self-confidence and personal motivation. *The Quarterly Journal of Economics* 117 (3): 871–915.

Benhabib, J., and A. Bisin. 2005. Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions. *Games and Economic Behavior* 52 (2): 460–492.

Berg, N., and G. Gigerenzer. 2010. As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*: 133–165.

Bernheim, B.D. (2009). The psychology and neurobiology of judgment and decision making: What's in it for economists?. In Glimcher, P. W., Camerer, C. F., Fehr, E., & Poldrack, R. A. (Eds.), *Neuroeconomics: Decision Making and the Brain* (pp. 113–125). Academic Press.

Bernheim, B.D., and A. Rangel. 2004. Addiction and cue-triggered decision processes. *The American Economic Review* 94 (5): 1558–1590.

Bernheim, B.D., and A. Rangel. 2007. Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review* 97 (2): 464–470.

Bonnefon, J.F. 2018. The pros and cons of identifying critical thinking with system 2 processing. *Topoi* 37 (1): 113–119.

Botvinick, M., and T. Braver. 2015. Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology* 66: 83–113.

Botvinick, M.M., and J.D. Cohen. 2014. The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science* 38 (6): 1249–1285.

Brocas, I., and J.D. Carrillo. 2008. The brain as a hierarchical organization. *The American Economic Review* 98 (4): 1312–1346.

Brocas, I., and J.D. Carrillo. 2014. Dual-process theories of decision-making: A selective survey. *Journal of Economic Psychology* 41: 45–54.

Buturovic, Z., and S. Tasic. 2015. Kahneman's failed revolution against economic orthodoxy. *Critical Review* 27 (2): 127–145.

Camerer, C. 1999. Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences* 96 (19): 10575–10577.

Camerer, C.F. 2007. Neuroeconomics: Using neuroscience to make economic predictions. *The Economic Journal* 117 (519): C26–C42.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donoghue, and M. Rabin. 2003. Regulation for conservatives: Behavioral economics and the case for" asymmetric paternalism". *University of Pennsylvania Law Review* 151 (3): 1211–1254.

Camerer, C., G. Loewenstein, and D. Prelec. 2005. Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 43 (1): 9–64.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford Press.

Colombo, M. 2014. Neural representationalism, the hard problem of content and vitiated verdicts. A reply to Hutto & Myin (2013). *Phenomenology and the Cognitive Sciences* 13 (2): 257–274.

D'Esposito, M., J.A. Detre, D.C. Alsop, R.K. Shin, S. Atlas, and M. Grossman. 1995. The neural basis of the central executive system of working memory. *Nature* 378: 279–281.

De Neys, W., ed. 2017. *Dual process theory 2.0*. Routledge.

De Neys, W., and J.F. Bonnefon. 2013. The 'whys' and 'whens' of individual differences in thinking biases. *Trends in Cognitive Sciences* 17 (4): 172–178.

Edwards, W. 1954. The theory of decision making. *Psychological Bulletin* 51 (4): 380.

Epstein, S. 1994. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist* 49 (8): 709–724.

Evans, J.S.B.T. (1989. Bias in human reasoning: Causes and consequences. *Lawrence Erlbaum Associates*, Inc.

Evans, J.S.B.T. (2006). Dual system theories of cognition: Some issues. *In Proceedings of the 28th Annual Meeting of the Cognitive Science Society,* 202–207.

Evans, J.S.B.T. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology* 59: 255–278.

Evans, J.S.B.T. (2009a). How many dual-process theories do we need? One, two, or many? In eds. Evans, J.S.B.T., & Frankish, K. (Eds.), *In two minds: Dual processes and Beyond* (pp. 33–54). Oxford university press.

Evans, J.S.B.T. (2009b). How many dual-process theories do we need? One, two, or many? In eds. Evans, J.S.B.T., & Frankish, K. (Eds.), *In two minds: Dual processes and Beyond* (pp. 33–54). Oxford University Press.

Evans, J.S.B.T. 2011. Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review* 31 (2–3): 86–102.

Evans, J.S.B.T., (2012). Dual process theories of deductive reasoning: facts and fallacies. In Holyoak, K., & Morrison, R. (Eds.), *The Oxford handbook of Thinking and Reasoning*, pp.115–133. Oxford University Press.

Evans, J.S.B.T. (2017). Dual process theory: Perspectives and problems. In De Neys, W. (Ed.), *Dual Process Theory 2.0* (pp. 145–164). Routledge.

Evans, J.S.B.T., and D.E. Over. 1996. *Rationality and reasoning*. Hove: Psychology Press.

Evans, J.S.B., and K. Frankish, eds. 2009. *In two minds: Dual processes and beyond*. Vol. 10. Oxford: Oxford University Press.

Evans, J.S.B.T., and K.E. Stanovich. 2013a. Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science* 8 (3): 223–241.

Evans, J.S.B.T., and K.E. Stanovich. 2013b. Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on Psychological Science* 8 (3): 263–271.

Evans, J.S.B.T., and P.C. Wason. 1976. Rationalization in a reasoning task. *British Journal of Psychology* 67 (4): 479–486.

Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. *The Delphi Report*.

Frederick, S. 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives* 19 (4): 25–42.

Fudenberg, D., and D.K. Levine. 2006. A dual-self model of impulse control. *American Economic Review* 96 (5): 1449–1476.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky.

Gigerenzer, G. (2004). Striking a blow for sanity in theories of rationality. In M. Augier & J. G.March (Eds.), *Models of a man: Essays in memory of Herbert A. Simon* (pp. 389–409). MIT Press.

Gigerenzer, G. 2007. *Gut feelings: The intelligence of the unconscious*. New York: Viking Press.

Gigerenzer, G. 2008. *Rationality for mortals*. New York: Oxford University Press.

Gigerenzer, G. 2015. On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology* 6 (3): 361–383.

Gigerenzer, G., and H. Brighton. 2009. Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science* 1 (1): 107–143.

Gigerenzer, G., and D.G. Goldstein. 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103 (4): 650–669.

Gigerenzer, G., and T. Regier. 1996. How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin* 119: 23–26.

Gigerenzer, G., and T. Sturm. 2012. How (far) can rationality be naturalized? *Synthese* 187 (1): 243–268.

Gigerenzer, G., P.M. Todd, and The ABC Research Group. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilbert, D.T. (1999). What the mind's not. In Chaiken, S., & Trope, Y. (Eds.), *Dual-process Theories in Social Psychology*, (pp.3–11). Guilford Press.

Gilbert, D.T. 2002. Inferential correction. In *Heuristics and biases: The psychology of intuitive judgment*, ed. T. Gilovich, D. Griffin, and D. Kahneman, 167–184. New York: Cambridge University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). Heuristics and biases*: The Psychology of Intuitive Judgment*. Cambridge university press.

Glimcher, P.W. 2003. The neurobiology of visual-saccadic decision making. *Annual Review of Neuroscience* 26 (1): 133–179.

Gollwitzer, P., Fujita, K., & Oettingen, G. (2004). Planning and the implementation of goals. In *Handbook of Self-regulation*: *Research, Theory, and Applications*. Guilford Press.

Grayot, J. 2019. From selves to systems: On the intrapersonal and intraneural dynamics of decision making. *Journal of Economic Methodology*: 1–20.

Grüne-Yanoff, T. 2017. Reflections on the 2017 Nobel memorial prize awarded to Richard Thaler. *Erasmus Journal for Philosophy and Economics* 10 (2): 61–75.

Harrison, G., and D. Ross. 2010. The methodologies of neuroeconomics. *Journal of Economic Methodology* 17 (2): 185–196.

Heilmann, C. (2016). Behavioral economics. In McIntyre, L., & Rosenberg, A. (Eds.), *The Routledge companion to philosophy of social science*. Taylor & Francis.

Heukelom, F. (2014). *Behavioral economics: A history.* Cambridge University Press.

Hutto, D. D., & Myin, E. (2013). *Radical enactivism: Basic minds without content*. MIT Press.

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. MIT Press.

Hutto, D.D., E. Myin, A. Peeters, and F. Zahnoun. 2018. Putting computation in its place. In *The Routledge handbook of the computational mind*, ed. M. Sprevak and M. Colombo, 272–282. London: Routledge.

Infante, G., G. Lecouteux, and R. Sugden. 2016. Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology* 23 (1): 1–25.

Kahneman, D. 2003a. Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review* 93 (5): 1449–1475.

Kahneman, D. 2003b. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist* 58 (9): 697–720.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In Kahneman, D., & Gilovich, T. (Eds.), *Heuristics and biases: The Psychology of Intuitive Judgment*, 49–81.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In Holyoak, K., & Morrison, R. (Eds.), *The Cambridge handbook of thinking and reasoning*, (pp. 267–293). Cambridge University Press.

Kahneman, D., and S. Frederick. 2007. Frames and brains: Elicitation and control of response tendencies. *Trends in Cognitive Sciences* 11 (2): 45–46.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263–292.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. *Judgement under uncertainty: Heuristics and biases*, (pp. 3–20). Cambridge University Press.

Keren, G. 2013. A tale of two systems: A scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science* 8 (3): 257–262.

Keren, G., and Y. Schul. 2009. Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science* 4 (6): 533–550.

Kirkpatrick, L.A., and S. Epstein. 1992. Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology* 63 (4): 534–544.

Konovalov, A., and I. Krajbich. 2019. Over a decade of neuroeconomics: What have we learned? *Organizational Research Methods* 22 (1): 148–173.

Krajbich, I., and M. Dean. 2015. How can neuroscience inform economics? *Current Opinion in Behavioral Sciences* 5: 51–57.

Kruglanski, A.W., and G. Gigerenzer. 2011. Intuitive and deliberative judgments are based on common principles. *Psychological Review* 118: 97–109.

Laibson, D. 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* 112 (2): 443–478.

Levin, J. (2018). "Functionalism", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/fall2018/entries/functionalism/.

Lieberman, M.D., Gaunt, R., Gilbert, D.T. and Trope, Y. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. In *Advances in experimental social psychology* (Vol. 34, pp. 199–249). Academic Press.

Loewenstein, G. 1996. Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes* 65 (3): 272–292.

Loewenstein, G. 2000. Emotions in economic theory and economic behavior. *The American Economic Review* 90 (2): 426–432.

Loewenstein, G., & O'Donoghue T. (2005). Animal spirits: Affective and deliberative processes in economic behavior. *CMU Working Paper*.

Lurquin, J.H., and A. Miyake. 2017. Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers in Psychology* 8: 568.

Mars, R., J. Sallet, M. Rushworth, and N. Yeung, eds. 2011. *Neural basis of motivational and cognitive control*. Cambridge: MIT Press.

McClure, S.M., D.I. Laibson, G. Loewenstein, and J.D. Cohen. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science* 306 (5695): 503–507.

McClure, S.M., K.M. Ericson, D.I. Laibson, G. Loewenstein, and J.D. Cohen. 2007. Time discounting for primary rewards. *Journal of Neuroscience* 27 (21): 5796–5804.

Mercier, H., and D. Sperber. 2011. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* 34 (2): 57–74.

Metcalfe, J., and W. Mischel. 1999. A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review* 106 (1): 3–19.

Miłkowski, M. (2013a). *Explaining the computational mind*. MIT Press.

Miłkowski, M. (2013b). Limits of computational explanation of cognition. In *Philosophy and theory of artificial intelligence* (pp. 69–84). Springer, Berlin, Heidelberg.

Mugg, J. 2016. The dual-process turn: How recent defenses of dual-process theories of reasoning fail. *Philosophical Psychology* 29 (2): 300–309.

Nisbett, R.E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall.

Osman, M. 2004. An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review* 11 (6): 988–1010.

Over, D. (2004). Rationality and the normative/descriptive distinction. In Koehler, D.J., & Harvey, N. (Eds.), *Blackwell handbook of judgment and decision making*, (pp. 3–18). John Wiley & Sons.

Parker, A.J., and W.T. Newsome. 1998. Sense and the single neuron: Probing the physiology of perception. *Annual Review of Neuroscience* 21 (1): 227–277.

Pennycook, G. (2017). A perspective on the theoretical foundation of dual process models. In De Neys, W. (Ed.), *Dual Process Theory 2.0* (pp. 13–35). Routledge.

Pennycook, G., J.A. Fugelsang, and D.J. Koehler. 2015. What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology* 80: 34–72.

Piccinini, G. 2015. *Physical computation: A mechanistic account*. Oxford: Oxford University Press.

Piccinini, G., and S. Bahar. 2013. Neural computation and the computational theory of cognition. *Cognitive Science* 37 (3): 453–488.

Samuels, R., & Stich, S. P. (2004). Rationality and psychology. The Oxford handbook of rationality, 279–300.

Sahlin, N.E., A. Wallin, and J. Persson. 2010. Decision science: From Ramsey to dual process theories. *Synthese* 172 (1): 129–143.
Samuels, R., Stich, S., & Bishop, M. (2012). Ending the rationality wars. *Collected Papers, Volume 2: Knowledge, Rationality, and Morality*, 1978–2010, 2, 191.
Sanfey, A.G., J.K. Rilling, J.A. Aronson, L.E. Nystrom, and J.D. Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300 (5626): 1755–1758.
Schelling, T.C. 1984. Self-command in practice, in policy, and in a theory of rational choice. *The American Economic Review* 74 (2): 1–11.
Sent, E.M. 2004. Behavioral economics: How psychology made its (limited) way back into economics. *History of Political Economy* 36 (4): 735–760.
Shefrin, H.M., and R.H. Thaler. 1988. The behavioral life-cycle hypothesis. *Economic Inquiry* 26 (4): 609–643.
Shiffrin, R.M., and W. Schneider. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review* 84 (2): 127–190.
Sinayev, A. (2016). Dual-System Theories of Decision Making: Analytic Approaches and Empirical Tests. (Electronic Thesis or Dissertation). Retrieved from https://etd.ohiolink.edu/. Accessed 19 Dec 2018.
Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin* 119 (1): 3.
Soman, D., G. Ainslie, S. Frederick, X. Li, J. Lynch, P. Moreau, A. Mitchell, D. Read, A. Sawyer, Y. Trope and K. Wertenbroch. 2005. The psychology of intertemporal discounting: Why are distant events valued differently from proximal ones?. *Marketing Letters* 16 (3–4): 347–360.
Stanovich, K.E. (1999). *Who is rational? Studies of individual differences in Reasoning*. Psychology Press.
Stanovich, K.E. (2004). *The Robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press.
Stanovich, K.E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. In eds. Evans, J.S.B.T., & Frankish, K. (Eds.), *In two minds: Dual processes and Beyond* (pp. 55–88). Oxford University Press.
Stanovich, K.E. (2011). *Rationality and the reflective mind*. Oxford University Press.
Stanovich, K.E., and R.F. West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23 (5): 645–665.
Stanovich, K. E., & West, R. F. (2009). *What intelligence tests Miss. The psychology of rational thought*. Yale University Press.
Stein, E. 1996. *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford: Oxford University Press.
Strack, F., and R. Deutsch. 2004. Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review* 8(3): 220–247.
Stroop, J.R. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18 (6): 643–662.
Swan, A.B., D.P. Calvillo, and R. Revlin. 2018. To detect or not to detect: A replication and extension of the three-stage model. *Acta Psychologica* 187: 54–65.
Szameitat, A.J., T. Schubert, K. Muller, and D.Y. von Cramon. 2002. Localization of executive function in dual-task performance with fMRI. *Journal of Cognitive Neuroscience* 14 (8): 1184–1199.
Thaler, R.H., and H.M. Shefrin. 1981. An economic theory of self-control. *The Journal of Political Economy* 89 (2): 392–406.
Thaler, R.H., and C.R. Sunstein. 2003. Libertarian paternalism. *American Economic Review* 93 (2): 175–179.
Thaler, R.H., & Sunstein, C.R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
Thaler, R.H., Sunstein, C.R., & Balz, J.P. (2012). Choice architecture. In Shafir, E. (Ed.), *The Behavioral Foundations of Public Policy*, (pp. 428–439). Princeton University Press.
Thompson, E. 2007. *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge: Harvard University Press.
Thompson, V. 2013. Why it matters: The implications of autonomous processes for dual process theories—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science* 8: 253–256.
Thompson, K.G., and J.D. Schall. 1999. The detection of visual signals by macaque frontal eye field during masking. *Nature Neuroscience* 2 (3): 283–288.
Tversky, A., and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5 (2): 207–232.
Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 (4157): 1124–1131.
Tversky, A., and D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5 (4): 297–323.
Tversky, A., & Kahneman, D. (Eds.). (2000). *Choices, values, and frames*. Cambridge University Press.

Van Gelder, T. 1995. What might cognition be, if not computation? *The Journal of Philosophy* 92 (7): 345–381.
Varga, A.L., and K. Hamburger. 2014. Beyond type 1 vs. type 2 processing: The tri-dimensional way. *Frontiers in Psychology* 5: 993.
Vromen, J. 2011. Neuroeconomics: Two camps gradually converging: What can economics gain from it? *International Review of Economics* 58 (3): 267–285.
Wakker, P. P. (2010). Prospect theory: For risk and ambiguity. Cambridge University Press.